Slope heuristics for multiple change-point models

Yann Guédon[1]

[1] CIRAD, UMR AGAP and Inria, Virtual Plants, Montpellier, France

E-mail for correspondence: guedon@cirad.fr

Abstract: With regard to multiple change-point models, much effort has been devoted to the selection of the number of change points. But, the proposed approaches are either dedicated to specific segment models or give unsatisfactory results for short or medium length sequences. We propose to apply the slope heuristic, a recently proposed non-asymptotic penalized likelihood criterion, for selecting the number of change points. In particular we apply the data-driven slope estimation method, the key point being to define a relevant penalty shape. The proposed approach is illustrated using two benchmark data sets.

Keywords: Data-driven slope estimation; Latent structure model; Model selection; Multiple change-point detection.

1 Introduction

The slope heuristics were introduced by Birgé and Massart (2001) as a new non-asymptotic penalized likelihood criterion for model selection. They showed that there exists a minimal penalty such that the dimension of models (and the associated estimator risk) selected with lighter penalties becomes very large. Moreover, they proved that considering a penalty equal to twice this minimal penalty allows to select a model close to the best possible (or oracle) model in terms of estimator risk. This approach has been recently popularized by the introduction of the data-driven slope estimation method by Baudry et al. (2012) which is a practical method for implementing slope heuristics. In the maximum likelihood estimation framework, this practical method is based on the expectation of a linear relation between the penalty shape (a function of the model dimension) and the maximized log-likelihoods for overparameterized models. We focus here on the application of the slope heuristics for selecting the number of change points in

This paper was published as a part of the proceedings of the 30th International Workshop on Statistical Modelling, Johannes Kepler Universität Linz, 6–10 July 2015.

multiple change-point models.

2 Defining the log-likelihood function and the penalty shape for multiple change-point models

For multiple change-point models, the two possible log-likelihood functions are:

- $\log f(\mathbf{s}^*, \mathbf{x}; J)$, the log-likelihood of the most probable segmentation $\mathbf{s}^*$ in J segments (the number of change points is therefore $J-1$) of the observed sequence $\mathbf{x}$. Lebarbier (2005) used this log-likelihood function for defining a slope heuristic for Gaussian models.

- $\log f(\mathbf{x}; J)$, the log-likelihood of all the possible segmentations in J segments of the observed sequence $\mathbf{x}$ with $f(\mathbf{x}; J) = \sum_{\mathbf{s}} f(\mathbf{s}, \mathbf{x}; J)$.

These log-likelihoods for $K = 2, \ldots, J$ can be exactly computed by a single application of a dynamic programming algorithm for $\log f(\mathbf{s}^*, \mathbf{x}; J)$ and a smoothing algorithm for $\log f(\mathbf{x}; J)$ (Guédon, 2013). The slope estimation relies on maximized log-likelihoods computed for overparameterized models and, as shown in Guédon (2013, 2015), the most probable segmentation is often meaningless for overparameterized models. Consistently with our view of multiple change-point models as latent structure models (Guédon, 2013; 2015), we will thus focus on the log-likelihood of all the possible segmentations $\log f(\mathbf{x}; J)$ for defining a slope heuristic. We investigated on several data sets this log-likelihood over the range of J values corresponding to overparameterized models and noted that this log-likelihood function is markedly concave for overparameterized models if $J < T$ (e.g. $10 < T/J < 100$), where T is the sequence length, but far less if $J \ll T$.
To apply the slope heuristics, it is required that (Baudry et al., 2012):
(C1) The log-likelihood increases with J.
(C2) The penalty shape $\mathrm{pen}_{\mathrm{shape}}(J)$ increases with J.
To these two standard requirements, we add the two following specific requirements for multiple change-point models:
(C3) The penalty shape $\mathrm{pen}_{\mathrm{shape}}(J)$ depends on the sequence length T. Adding a segment for say $J = T/10$ entails a smaller increase of the penalty shape than adding a segment for $J \ll T$.
(C4) The first-order differenced penalty shape $\mathrm{pen}_{\mathrm{shape}}(J) - \mathrm{pen}_{\mathrm{shape}}(J-1)$ decreases with J. This decrease is not a linear function of J.
For the definition of the penalty shape, our starting point was $\mathrm{pen}_{\mathrm{shape}}(J) = \log n_J$ where $n_J = \binom{T-1}{J-1}$ is the number of possible segmentations in J segments. Consider the limiting case where all the segmentations are equally probable for a fixed J, then $\log f(\mathbf{x}; J) = \log \gamma_J + \log n_J$. In fact for overparameterized models, as shown in Guédon (2013, 2015) using different examples, $\log f(\mathbf{x}; J)$ decomposes into a structural part corresponding to

true change points and a noise part that increases with J. To respect the monotonicity of $\mathrm{pen}_{\mathrm{shape}}(J)$ as a function of J, we finally propose

$$\mathrm{pen}_{\mathrm{shape}}(J) = \log\left\{\frac{T^{J-1}}{(J-1)!}\right\},$$

with

$$\mathrm{pen}_{\mathrm{shape}}(J) - \mathrm{pen}_{\mathrm{shape}}(J-1) = \log\left(\frac{T}{J-1}\right).$$

3 Illustrations on benchmark data sets

The use of the proposed slope heuristic for multiple change-point models is illustrated using two benchmark data sets corresponding to different segment models and sequence lengths. The slope heuristic is compared with the “exact”ICL criterion proposed by Rigaill et al. (2012).

3.1 British coal mining disasters

The data consist of the dates of 191 coal mining disasters between 1851 and 1962 summarized as annual counts during the 112-year period. We assume that the number of disasters in any year has a Poisson distribution, and the underlying Poisson distribution parameter is piecewise constant through time.

TABLE 1. British coal mining disaster data: comparison of the exact ICL criterion and the slope heuristics (SH) with J as the penalty shape ($\mathrm{pen}_{\mathrm{shape}}0$) and the proposed penalty shape ($\mathrm{pen}_{\mathrm{shape}}1$). The criterion value and the corresponding posterior model probability $P(\mathcal{M}_J|\mathbf{x})$ are given for each J.

J	1	2	3	4	5
ICL_J	−413.14	−358.70	−362.31	−369.34	−375.74
$P(\mathcal{M}_J\|\mathbf{x})$	0	0.855	0.141	0.004	0
SH_J ($\mathrm{pen}_{\mathrm{shape}}0$)	−419.68	−358.81	−357.04	−358.55	−361.01
$P(\mathcal{M}_J\|\mathbf{x})$	0	0.2	0.49	0.23	0.07
SH_J ($\mathrm{pen}_{\mathrm{shape}}1$)	−407.72	−360.19	−368.05	−377.00	−385.38
$P(\mathcal{M}_J\|\mathbf{x})$	0	0.98	0.02	0	0

The slopes were estimated over the range $J = 6, \ldots, 20$ and, we obtained a residual standard deviation of 1.05 with the naive penalty shape J instead of 0.04 with the proposed penalty shape. Both the exact ICL criterion and the slope heuristic with the proposed penalty shape favour 2 segments while the slope heuristic with the naive penalty shape J favours 3 segments and puts weight on 4 segments which is not consistent with the outputs of the different validation approaches shown in Guédon (2013, 2015); see Table 1.

3.2 Well-log data

The data consist of 4050 measurements of the nuclear-magnetic response of underground rocks. The underlying signal is roughly piecewise constant, with each segment relating to a single rock type that has constant physical properties. We estimated Gaussian change in the mean and variance models on the basis of these data.

TABLE 2. Well-log data: comparison of the exact ICL criterion and the slope heuristic (SH) with the proposed penalty shape. The criterion value and the corresponding posterior model probability $P(\mathcal{M}_J|\mathbf{x})$ are given for each J.

J	15	16	17	18	19	20
ICL_J	−69355.4	−69330.0	−69316.3	−69309.7	−69309.6	−69313.3
$P(\mathcal{M}_J\|\mathbf{x})$	0	0	0.014	0.378	0.403	0.063
SH_J	−69479.6	−69461.8	−69466.0	−69478.6	−69482.8	−69495.4
$P(\mathcal{M}_J\|\mathbf{x})$	0	0.89	0.11	0	0	0

The slope was estimated over the range $J = 30, \ldots, 80$. The exact ICL criterion favours 18 and 19 segments while the slope heuristic with the proposed penalty shape favours mainly 16 segments; see Table 2. The 16-segment model selected by the slope heuristic is far more consistent with the analysis of the segmentation space presented in Guédon (2013) than the 18- or 19-segment models selected by the exact ICL criterion.

References

Baudry, J.-P., Maugis, C., and Michel, B. (2012). Slope heuristics: overview and implementation. *Statistics and Computing*, **22**, 455 – 470.

Birgé, L. and Massart, P. (2001). Gaussian model selection. *Journal of the European Mathematical Society*, **3**, 203 – 268.

Guédon, Y. (2013). Exploring the latent segmentation space for the assessment of multiple change-point models. *Computational Statistics*, **28**, 2641 – 2678.

Guédon, Y. (2015). Segmentation uncertainty in multiple change-point models. *Statistics and Computing*, **25**, 303 – 320.

Lebarbier, E. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing*, **85**, 717 – 736.

Rigaill, G., Lebarbier, E., and Robin, S. (2012). Exact posterior distributions and model selection criteria for multiple change-point detection problems. *Statistics and Computing*, **22** 917 – 929.